

3D Genome Tuner: Compare Multiple Circular Genomes in a 3D Context

Qi Wang^{1,2*}, Qun Liang^{1,2}, and Xiuqing Zhang^{1,3}

¹Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China; ²Graduate University of Chinese Academy of Sciences, Beijing 100049, China; ³Beijing Genomics Institute at Shenzhen, Shenzhen 518083, China.

*Corresponding author. E-mail: qwang.big@gmail.com

DOI: 10.1016/S1672-0229(08)60043-1

Circular genomes, being the largest proportion of sequenced genomes, play an important role in genome analysis. However, traditional 2D circular map only provides an overview and annotations of genome but does not offer feature-based comparison. For remedying these shortcomings, we developed 3D Genome Tuner, a hybrid of circular map and comparative map tools. Its capability of viewing comparisons between multiple circular maps in a 3D space offers great benefits to the study of comparative genomics. The program is freely available (under an LGPL licence) at <http://sourceforge.net/projects/dgenometuner>.

Key words: three dimensional, circular map, genome browser, comparative genomics

Introduction

Advances in whole-genome sequencing technology have resulted in massive amounts of genomic data. The best way to learn the characteristics of a genome and genetic mutation between genomes is using visualization applications or so-called genome browsers. **Table 1** lists a comparison between some browsers. As we can see, modern browsers are generally platform-independent, rich in presentation methods and reading common sequence formats. Furthermore, it is noticeable that linear sequence browsers are usually used in comparisons while circular sequence browsers are used in presentations, mainly because of the absence of space to accommodate multiple genomes in a circular map, and the inconvenience to draw links between circular genomes. Though comparisons can be taken in linear form for circular genomes, the lines linking the head of one genome and the tail of another are remarkably stretched (**Figure 1A**). However, the same links are far less affected in a circular comparison map (**Figure 1B**). Recently, we discovered that it is more elegant to have the circles placed in a 3D space, as in this mode it is possible to accommodate more information for each circle and it can be easily viewed by rotating one of the circles (**Figure 1C**).

Taking into consideration that most of the completed sequences, typically short or prokaryotic ones, are circular DNA (bacterial, plasmid, mitochondrial

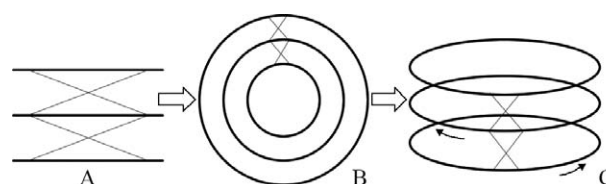


Figure 1 From linear view (**A**) to 2D circular view (**B**) then 3D circular view (**C**). The lines linking homologous regions are condensed to offer a clearer comparison.

and chloroplast), we developed a new program, called 3D Genome Tuner, which enables comparing circular genomes in a 3D space.

Application

Implementation

3D Genome Tuner is a stand-alone application employing Java Swing and JOGL API (JSR-231 1.1.1). Some open source libraries such as Jaligner and Jfreechart are incorporated in the package. It adopts MVC (Model-View-Controller) architecture and has been tested under Microsoft Windows and Fedora Linux. The program is freely available (under an LGPL licence) at <http://sourceforge.net/projects/dgenometuner>.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1 Comparison of features of some freely available, stand-alone genome browsers

Program (Reference)	Platform ^{*1}	Input format	Circular view	Linear view	Dot view	Genome comparison ^{*2}	Dimensional
3D Genome Tuner	Java	EMBL, GBK, FASTA, BLAST	✓	✓		✓	2D/3D
ACT (7)	Java	EMBL, GBK, FASTA, GFF		✓		✓	2D
Bluejay (8)	Java	XML	✓	✓		✓	2D
BugView (9)	Java	GBK		✓		✓ (two)	2D
CGAT (10)	Java (client), Perl (server)	GBK, FASTA		✓	✓	✓ (two)	2D
CGView (11)	Java	PTT, XML	✓				2D
Combo (12)	Java	GBK, FASTA, GFF, GENSCAN		✓	✓	✓ (two)	2D
DNAPlotter (13)	Java	EMBL, GBK, FASTA, GFF	✓	✓			2D
DNAMVis (14)	Windows, Linux	GFF, FASTA		✓	✓	✓	2D/3D
GATA (15)	Java	GFF		✓		✓ (two)	2D
GeneVTio (16)	Java	PTT, FNN	✓	✓			2D
GenoMap (17)	Tcl/Tk	GTF	✓				2D
Genome2D (18)	Windows	tab delimited		✓		✓	2D
GenomeComp (19)	Perl/Tk	EMBL, GBK, FASTA		✓		✓ (two)	2D
GenomeMatcher (20)	Mac	GBK, FASTA		✓	✓	✓ (two)	2D
GenomePlot (21)	Perl/Tk	tab delimited	✓	✓			2D
GenomeViz (22)	Tcl/Tk/Perl	tab delimited	✓			✓	2D
Genomorama (23)	Mac, Linux, Windows	EMBL, GBK, FASTA, PTT, ASN.1		✓		✓	2D
Mauve (3)	Java	GBK, SEQ, FASTA		✓		✓	2D
Microbial Genome Viewer (24)	online	EMBL, GBK, tab delimited	✓	✓			2D
MUMmerplot (25)	Linux	FASTA			✓	✓ (two)	2D
SeqVISTA (26)	Java	GBK, FASTA		✓		✓	2D
Sockeye (27)	Java	EMBL, GFF		✓		✓	3D

^{*1}Programs developed using Java, Tcl/Tk or Perl are considered to be platform-independent. ^{*2}Software with “✓ (two)” means that it can only compare two genomes at one time, while those with “✓” have no limits.

User interface

3D Genome Tuner is an interactive application and provides real-time visualization (**Figure 2**). With rotate and zoom buttons, users can easily move camera or scale objects. By selecting one map to rotate while leaving others fixed in place, maps can be aligned within features. 3D Genome Tuner draws on a fixed size of screen and resolves very closely spaced features by zooming in and out. The number of polygons may affect its performance. However, there is no lag in the range of prokaryotic genome size (1). During the test, we found that even the largest bacterial chromosome longer than 8,000,000 bp can be loaded and displayed comfortably on a computer with 512 Mb of memory and a moderate 3D graphics card. 3D Genome Tuner gives user privilege to set the color of any item and the radius of any track on the map. Once changes have been made, the settings are stored in a properties file

under the same folder as the program, which will be loaded at the next startup. The graph can also be exported to an image for publication or demonstration purpose.

Data acquisition

3D Genome Tuner accepts sequences in EMBL, GenBank or FASTA format. Files can be accessed from a local disk or network. The program has recorded the URLs of all bacterial and other circular genomes on NCBI ftp server, so users only need to select a genome and download from the bookmarks. Unlimited number of sequences can be opened at one time and drawn in a stack. Once a sequence is loaded, three circles are generated by default: GC-content, GC-skew and the coordinates. Genes can either be recognized from GenBank or EMBL annotations, or by loading tab-

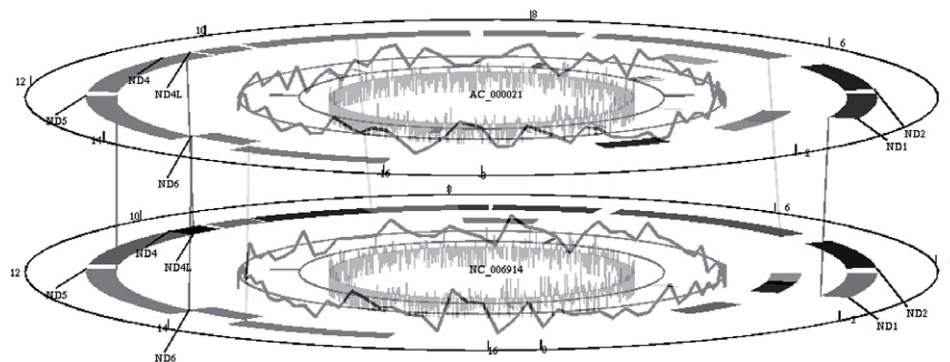


Figure 2 A screenshot showing comparison between two mitochondrion genomes from *Homo sapiens* (AC_000021) and *Mus musculus domesticus* (NC_006914). The tracks from the outside to inside represent: (1) coordinates; (2) CDSs; (3) homologous regions calculated by NUCmer; (4) GC-content; (5) tRNA and rRNA genes; (6) GC-skew $[(G-C)/(G+C)]$. Lines linking the upper and lower circular maps are BLAST alignments between the orthologous genes.

delimited files containing their locations. Genomic regions specified by user or imported from MUMmer (2) or Mauve (3) outputs may also take up a circle. To show comparison between genes, BLAST (4) alignment result is required. 3D Genome Tuner summarizes redundant BLAST matches by choosing the one with longest overlap among these matches. Duplicated matches with the same score are also reserved. The orthologous genes are linked with different colors indicating the matching percentage. MUMmer or Mauve output describing homologous regions between genomes can be parsed and appended to maps in the same manner, except that regions are painted with colors based on their orders.

3D Genome Tuner allows user to choose different classification methods to render CDS (coding sequence) genes, which helps them better understand the genome organizations. By providing codes in Clusters of Orthologous Groups of proteins (COGs) (5) or Gene Ontology (GO) (6), genes with the same biological function are painted with the same colors. Other classification methods are based on the characteristics of the genes, such as length or GC-content. By dividing genes into several groups and using gradient colors in these groups, the distribution of genes can be manifested at a glance.

Exploration

3D Genome Tuner provides one navigation window and five sub-windows that allow user to explore detailed genomic features. The navigation window is a tree control that expands on every genome as well as every feature. The items in the tree control provide tool tips regarding the properties of the genome or genes, *etc.* Click an item representing a CDS, RNA

or pseudo gene, then the locus tag of the gene will be labelled besides its location in the graph. If an orthologous gene pair is clicked, a line linking the location of each gene is shown. Below the main window, the first sub-window shows a linear image of current selected genome, in which the CDS and RNA genes are drawn in lines with arrowheads representing their strands. The second sub-window is a plot showing frame-specific statistics, including GC-content, GC-skew and gene coverage. The third sub-window is a sequence browser in which nucleotides will be shown when the item for corresponding gene is clicked in the navigation window. The fourth sub-window shows alignment between amino acids of the orthologous gene pair that user clicked. The alignment is performed by Jaligner (<http://jaligner.sourceforge.net>) at real time. Parameters for the alignment can be set in the option panel. The last sub-window shows a zoom-in picture of the main canvas.

3D Genome Tuner carries out simple analysis by highlighting genomic regions of interests concerning high variance of GC-content, length or count of genes and gene coverage. The analysis assumes that the values of every frame are in normal distribution. Values that fall out of 95% or 99% confidence interval are drawn in pies filled with corresponding colors on the map, aiding researchers in further investigations.

Conclusion

3D Genome Tuner shows the power of 3D graphics in genome studies. It is a novel tool that enables viewing multiple circular genomes in a 3D context by presenting information in distinct dimensions. The improvements include: (1) genomes can be compared

in circular form; (2) alignments won't be stretched too long. 3D Genome Tuner is a tool for preliminary genome studies and provides a possible way to resolve the difficulties in the comparison between circular genomes. Its ease-of-use not only makes the research more effective but also brings user a delightful experience in exploring with animation and sprite colors. The future version of 3D Genome Tuner will focus on the integration of other analysis tools and provide greater usability to users. XML annotations and microarray data support are both on schedule. The graphics will also be updated to provide more precise features and annotations. We hope 3D Genome Tuner will open a new perspective in genome visualizing researches and facilitate studies of comparative genomics.

Acknowledgements

This project was supported by the National High-Tech R&D Program (863 Program) of China (Grant No. 2006AA02A301) awarded to XZ.

Authors' contributions

QW designed and implemented the program as well as wrote and revised the manuscript. QL participated in drafting the manuscript and testing the program. XZ supervised the study and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Doolittle, R.F. 2002. Biodiversity: microbial genomes multiply. *Nature* 416: 697-700.
2. Delcher, A.L., *et al.* 1999. Alignment of whole genomes. *Nucleic Acids Res.* 27: 2369-2376.
3. Darling, A.C., *et al.* 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14: 1394-1403.
4. Altschul, S.F., *et al.* 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
5. Tatusov, R.L., *et al.* 1997. A genomic perspective on protein families. *Science* 278: 631-637.
6. Ashburner, M., *et al.* 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25: 25-29.
7. Carver, T.J., *et al.* 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21: 3422-3423.
8. Soh, J., *et al.* 2008. Bluejay 1.0: genome browsing and comparison with rich customization provision and dynamic resource linking. *BMC Bioinformatics* 9: 450.
9. Leader, D.P. 2004. BugView: a browser for comparing genomes. *Bioinformatics* 20: 129-130.
10. Uchiyama, I., *et al.* 2006. CGAT: a comparative genome analysis tool for visualizing alignments in the analysis of complex evolutionary changes between closely related genomes. *BMC Bioinformatics* 7: 472.
11. Stothard, P. and Wishart, D.S. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21: 537-539.
12. Engels, R., *et al.* 2006. Combo: a whole genome comparative browser. *Bioinformatics* 22: 1782-1783.
13. Carver, T., *et al.* 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25: 119-120.
14. Fiers, M.W., *et al.* 2006. DNAVis: interactive visualization of comparative genome annotations. *Bioinformatics* 22: 354-355.
15. Nix, D.A. and Eisen, M.E. 2005. GATA: a graphic alignment tool for comparative sequence analysis. *BMC Bioinformatics* 6: 9.
16. Vernikos, G.S., *et al.* 2003. GeneViTo: visualizing gene-product functional and structural features in genomic datasets. *BMC Bioinformatics* 4: 53.
17. Sato, N. and Ehira, S. 2003. GenoMap, a circular genome data viewer. *Bioinformatics* 19: 1583-1584.
18. Baerends, R.J., *et al.* 2004. Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data. *Genome Biol.* 5: R37.
19. Yang, J., *et al.* 2003. GenomeComp: a visualization tool for microbial genome comparison. *J. Microbiol. Methods* 54: 423-426.
20. Ohtsubo, Y., *et al.* 2008. GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinformatics* 9: 376.
21. Gibson, R. and Smith, D.R. 2003. Genome visualization made fast and simple. *Bioinformatics* 19: 1449-1450.
22. Ghai, R., *et al.* 2004. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* 5: 198.
23. Gans, J.D. and Wolinsky, M. 2007. Genomorama: genome visualization and analysis. *BMC Bioinformatics* 8: 204.
24. Kerkhoven, R., *et al.* 2004. Visualization for genomics: the Microbial Genome Viewer. *Bioinformatics* 20: 1812-1814.
25. Kurtz, S., *et al.* 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5: R12.
26. Hu, Z., *et al.* 2003. SeqVISTA: a graphical tool for sequence feature visualization and comparison. *BMC Bioinformatics* 4: 1.
27. Montgomery, S.B., *et al.* 2004. Sockeye: a 3D environment for comparative genomics. *Genome Res.* 14: 956-962.